

AD-A008 410

DOMINANCE OF INDIVIDUATING INFORMATION
AND NEGLECT OF BASE RATES IN
PROBABILITY ESTIMATION

Don Lyon, et al

Oregon Research Institute

Prepared for:

Office of Naval Research

21 February 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

ADA008410

119106

ONR Technical Report

DOMINANCE OF INDIVIDUATING INFORMATION AND NEGLECT OF BASE RATES IN PROBABILITY ESTIMATION

Don Lyon
Paul Slovic



This research was supported by the Advanced Research Projects Agency of the Department of Defense (ARPA Order No. 2449) and was monitored by ONR under Contract No. N00014-73-C-0438 (NR 197-026).

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

Reproduction in whole or part is permitted for any purpose of the United States Government.

This report also appears as Oregon Research Institute Research Bulletin Vol. 15, No. 1.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield, VA 22151

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

Unclassified

Security Classification

AD-A 008 410

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Oregon Research Institute Eugene, Oregon		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE Dominance of Individuating Information and Neglect of Base Rates in Probability Estimation			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report			
5. AUTHOR(S) (First name, middle initial, last name) Don Lyon Paul Slovic			
6. REPORT DATE February 21, 1975		7a. TOTAL NO. OF PAGES 21	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. N00014-73-C-0438		9a. ORIGINATOR'S REPORT NUMBER(S) Oregon Research Institute Research Technical Report Vol. , No.	
b. PROJECT NO. NR(197--026)			
c. ARPA Order No. 2449		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.			
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale. Its distribution is unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Office of Naval Research Code 455 Arlington, Virginia 22217	
13. ABSTRACT According to the normative theory of prediction, prior probabilities (base rates) which summarize what we know before receiving any specific evidence, should remain relevant even after such evidence is obtained. In the present study, subjects were asked to estimate the probability that one of two states was true on the basis of (a) information about the prior probabilities of the states and (b) individuating information, specific to the case at hand and known to be accurate with probability p. Subjects' responses were determined predominantly by the specific evidence; the prior probabilities were neglected, causing the judgments to deviate markedly from the normative response. Theoretical and practical implications of this result are discussed.			

DDC
RECEIVED
APR 17 1975
RECEIVED
D

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

DD FORM 1473

REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Security Classification

1

Dominance of Individuating Information and
Neglect of Base Rates in Probability Estimation

Don Lyon
University of Oregon

Paul Slovic
Oregon Research Institute

Running Head: Dominance of Individuating Information

A fundamental principle of the normative theory of prediction is that prior probability, which summarizes what we know before receiving any specific evidence, remains relevant even after such evidence is obtained. Recent investigations by Kahneman and Tversky (1973) and Hammerton (1973) have described common prediction situations in which people deviate from the normative prescription by relying almost exclusively on specific evidence and neglecting prior considerations. Kahneman and Tversky noted that their subjects' failure to appreciate the relevance of prior probability in the presence of specific information was "one of the most significant departures of intuition from the normative theory of prediction" (p. 243).

In the experiments by Kahneman and Tversky, the subjects' task was to estimate the probability that a target person belonged to a given academic or occupational group. Information specific to the case, henceforth referred to as "individuating information," was provided in the form of a short personality description such as: "John is a 45-year old man. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his hobbies which include carpentry, sailing, and mathematical puzzles." In this case, the prior probability, or base rate, was varied by telling subjects that the person described was randomly selected from a group consisting of either 30 engineers and 70 lawyers or, in another condition 70 engineers and 30 lawyers. Manipulation of the lawyer-engineer base rate was found to have little effect on subjects' estimates of the probability that the description was that of an engineer. Kahneman and Tversky argued that these probability estimates depended, instead, upon the degree of similarity between the description and the subjects' stereotypes for engineers

and lawyers. Kahneman and Tversky further noted that predicting outcomes according to their similarity to the individuating evidence reflects a general heuristic principle that people use when making probabilistic judgments. This heuristic, which they called "representativeness," leads to a variety of systematic errors and, in this task, caused judgments to neglect prior probabilities.

Hammerton independently found a similar result. He presented subjects with six statements such as the following:

- (1) A device has been invented to test for Disease X.
- (2) The device is a very good one, but not perfect.
- (3) If someone suffers from X, there is a 90% chance he will test positively.
- (4) If he is not a sufferer, there is still a 1% chance that he will be recorded positively.
- (5) Roughly 1% of the population has the disease.
- (6) Mr. Smith has been tested and the result is positive.

Hammerton's subjects were then asked to indicate the probability that Mr. Smith has Disease X. The probability of Disease X, given a positive test result, can be computed from Bayes' Theorem on the basis of the information in Statements 3, 4, and 5. The Bayesian probability turns out to be .48. Subjects greatly overestimated; their median probability was .85. In case the reader's perceptions follow the subjects', the veracity of the Bayesian calculation is illustrated by the 2 X 2 diagram in Table 1, showing the expected results from 10,000 test cases.

Insert Table 1 about here

The judgments of Hammerton's subjects seemed to be dominated by the conditional probability in the upper row of Table 1, rather than the conditional probability in the left column. The subjects' misunderstanding was further demonstrated when Hammerton removed various combinations of Statements 3, 4, and 5 from the problem description, thus rendering the problem unsolvable, but left one item of individuating information (Statement 2). Subjects still unanswered with median probabilities between .75 and .90. Furthermore, these answers were given with a high degree of confidence!

Hammerton's results, showing that subjects' estimates were dominated by the diagnosticity of the individuating data, are similar to those of Kahneman and Tversky. However, Hammerton interpreted them somewhat differently. He argued that subjects entered the experiment with a "rigid prior" expectation that diagnostic tests are infallible. He attempted to support this conjecture in a second study in which the content of the problem was changed to reduce the likelihood of subjects having strong prior expectations about diagnostic infallibility. Statement 1 was changed to read: "A device has been invented for screening engine parts for internal cracks." The other statements were altered to match, but the probabilities in the statements remained the same. Subjects' median probability that an engine with a positive test result was cracked was .60--significantly lower than the .85 typical of the medical problem, yet still higher than the Bayesian answer. Hammerton concluded that, although subjects still underweighted the base rate, this effect of changing content supported his hypothesis about "rigid priors."

Hammerton's results suggest that the dominance of individuating information may be affected by the content of the problem. Further investigation

of content seems warranted, and, in this regard, a pilot study done by Kahneman and Tversky (Note 1) is relevant. They gave subjects the following problem:

Two cab companies, the Blue and the Green, operate in a given city. Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit and run accident at night.

A witness identified the cab as a Green cab. The court tested his ability to distinguish a Blue cab from a Green cab at night, and found that he was able to make correct identifications in four out of five tries.

What do you think is the probability (expressed as a percentage) that the cab involved in this accident was Blue?

The median estimated probability that the cab was Blue was .20, indicating that, here too, subjects ignored base rate information and relied on the individuating information about the witness' accuracy. This study, and Hammerton's, suggest that judgments can be dominated by individuating information in numerical form as well as by the verbal descriptions in Kahneman and Tversky's other problems. But the presence of this individuation effect in the Cab Problem, where no "rigid prior" about witness' infallibility would be expected, brings Hammerton's explanation into question.

At this point, delineation of the conditions under which prior information is used or discarded seems to merit further investigation. The present study employed Hammerton's problems and several variations of the Cab Problem to test hypotheses about variables that could mitigate the effect. Specifically, problem content was varied. In addition, base rates and individuating accuracies were expressed as percentages in some prob-

lems and small numbers in others. Extreme base rates were studied. Random sampling from the population in question was emphasized. Individuating evidence that was congruent with the base rate was studied, as well as evidence in opposition to the base rate. Order of presentation and response format were also varied.

Neglect of prior probabilities in all these circumstances would have considerable theoretical and practical importance. Theoretical import stems from the role these findings would play in evaluating Hammerton's views about content and Kahneman and Tversky's hypothesis about representativeness as a determiner of subjective probability. Practical benefits would accrue from the recognition and correction of a systematic error which has potentially serious consequences for decision making. As Kahneman and Tversky state with regard to the Cab Problem: "Much as we would like to, we have no reason to believe that the typical juror does not evaluate evidence in this fashion" (Note 1, p. 13).

Method

The experiment employed numerous variations of the Cab Problem. The basic problem was modified to read as follows:

Two cab companies, the Blue and the Green, operate in a given city. Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. A witness identified the cab as a Green cab.

The court tested the witness' ability to distinguish a Blue cab from a Green cab at night by presenting to him film sequences, half of which depicted Blue cabs, and half depicting Green cabs. He was able to make correct identification in 8 out of 10 tries. He made one error on each color of cab.

What do you think is the probability (expressed as a percentage) that the cab involved in this accident was Blue?

This modification of the original problem was designed to specify that the witness' accuracy was estimated by the court test to be .8 regardless of the true color of the cab. Without specifying the witness' accuracy conditional upon each of the colors being true, the problem does not have a unique Bayesian solution. With it, the Bayesian probability that the cab is Blue, given the witness says Green, is .59 (see Table 2).

Insert Table 2 about here

Variations

Incidental features. Two incidental features of the Cab Problem were changed to establish whether they were influencing the judgments. Order of presentation was varied by giving the individuating information before the base rate information. The question in the standard problem was also changed to ask for the probability that the cab involved was Green.

Confirming evidence. In the standard Cab Problem, the witness' testimony goes counter to the base rate. Two versions of the problem were studied in which the witness claimed that the cab was Blue, thus confirming the implication of the base rate. In one version, the base rate was kept at 85% Blue, but the witness' accuracy was lowered to 60%. In the other version the base rate was lowered to 60% Blue, but the witness' accuracy was 80%.

Base rate alone. Kahneman and Tversky reported that, when individuating information was absent, subjects correctly used the base rate. To verify this result, one version of the Cab Problem had the witness' statement deleted.

Content. Another problem was created, similar to the Cab Problem in structure but not content. The "Light-Bulb Problem" takes the situation out of a court of law, employs a mechanical "witness," emphasizes random selection from the population, and presents the individuating information as the base rate is given, i.e., as a percentage. It read as follows:

A light bulb factory uses a scanning device which is supposed to put a mark on each defective bulb it spots in the assembly line. Eighty-five percent of the light bulbs on the line are OK; the remaining 15% are defective.

When a bulb is good, the scanner correctly identifies it as good 80% of the time. When a bulb is bad, the scanner correctly marks it 80% of the time.

Suppose someone selects one of the light bulbs from the line at random and gives it to the scanner. The scanner marks this bulb as defective.

What do you think is the probability (expressed as a percentage) that this bulb is really OK?

One version of this problem presented the individuating information in the same manner as the Cab Problem, saying that the scanner was able to correctly identify the condition of the bulb in eight out of ten trials, and it made one error on each kind of bulb. A second version of this problem had the phrase "at random" deleted.

As a further test of content effects, Hammerton's Disease Problem and Engine-Crack Problem were also studied here.

Base rate extremity. Several additional versions of the Light-Bulb Problem were studied. The base rate of defective bulbs was made 1 in 100,

instead of 15 in 100, and one group estimated the probability that the bulb was good while a second group estimated the probability that it was defective, given individuating evidence calling it defective.

Base rate format. Another Light-Bulb problem attempted to individuate the base rate by presenting it in a form similar to the presentation of individuating information in the standard Cab Problem. The crucial sentence read: "Suppose there are 10 bulbs to be tested. Eight of these are O.K. The other two are defective."

Base rate salience. Another problem was developed to test the hypothesis that well established or easily remembered (i.e., salient) base rates might be incorporated into the relevant impression and thus be given more weight in the estimate.

The 'Right-Hander' Problem began as follows:

A knifing incident was recently the subject of a jury trial in a particular city. A central issue in the case was whether the assailant was right-handed or left-handed. About 85% of the population of the city is right-handed; the remaining 15% are left-handed.

A witness testified in court that the attacker had held the knife in his left hand.

The remainder of the problem paralleled the structure for the other problems, with the witness being shown, by court test, to be 80% accurate in his ability to determine both left and right handed assailants under circumstances similar to those in question. The subject was then asked to estimate the probability that the assailant was right handed.

Response format. Each of the variations described above asked the subject to answer in terms of a probability or a percentage. To provide

a broader perspective on subjects' attitudes toward the relevance of base rates and individuating information, the basic Cab Problem and the Light-Bulb Problem were also studied with multiple-choice formats in which three opposing positions were stated. For example, in the Cab Problem, subjects were asked:

What do you think is the probability (expressed as a percentage) that the cab involved in this accident was Blue?

Indicate (by marking with an X) which of the following three answers best represents your opinion.

_____ a. Since the witness was 80% accurate in the court test, the probability that the cab which committed this crime is Blue is about 20%. The proportion of Blue and Green cabs in the city is irrelevant because it says nothing about the particular cab involved.

_____ b. Since 85% of the cabs in the city are Blue the probability that the cab which committed this crime is Blue is about 85%.

_____ c. The witness' statement and the proportion of Blue and Green cabs in the city are equally important. Since these two items of information point toward different conclusions, they offset one another. Therefore, the probability that the cab which committed this crime is Blue is about 50%.

How confident are you that your answer is appropriate? Mark one response.

1
Not at all
confident

2

3
Moderately
Confident

4

5
Very
Confident

Subjects

The basic Cab Problem, its variations, and the two Hammerton problems were presented in group settings to a total of 310 volunteers enrolled in psychology classes at the University of Oregon, and to 242 paid subjects at the Oregon Research Institute. Each subject saw only one problem, in written form.

Results

Since there was no discernable difference between the results from paid and volunteer subjects, data from both types of subjects were combined and are shown in Table 3. The standard version of the Cab Problem (line 1) produced results similar to those obtained by Kahneman and Tversky. The median and modal estimate was 20% probability of Blue, thus confirming the tendency of subjects to neglect the base rate and rely predominantly on the witness. Placing the individuating information first (line 2) or asking subjects to estimate the probability that the cab was Green (line 3) had no significant effect on the dominance of the individuating information. The results shown on lines 4 and 5 indicate that individuating evidence which confirmed the base rate was equally dominant. When the confirming witness was said to be 60% accurate, subjects tended to estimate the probability of Blue at 60%, and when the confirming witness was termed 80% accurate, subjects median estimate was 80%.

Insert Table 3 about here

As in the experiments by Kahneman and Tversky, the present subjects did use the base rate information when no individuating evidence was provided (line 6).

Turning to the Light Bulb Problems, the change in cover story (line 7), emphasis on random selection (line 8), and change in the base rate format (line 9) all had little or no effect. The median estimates for each problem matched the individuating accuracies exactly. Describing the individuating accuracy as eight out of ten correct, rather than 80% correct, also made no difference in the results.

When subjects estimated the probability that the bulb was good, the problem with the extreme base rate (line 10) produced a somewhat higher median probability estimate than did the comparable problem with the less extreme base rate (line 8). However, the difference between medians was not statistically significant ($\chi^2 = 2.12$; $df = 1$; $p > .05$). Furthermore, when subjects estimated the probability that the bulb was defective (line 11) they followed the individuating information very closely despite the extreme base rate.

The salience manipulation (line 12) failed to alter subjects' reliance on the individuating information. The responses to both of Hammerton's problems (lines 13 and 14) were dominated by the individuating information, leading subjects' estimates to exceed greatly the Bayesian answer of .48. In contrast to Hammerton's results, there was no significant tendency in the present data for the Disease Problem to elicit higher probability estimates than the Engine-Crack Problem.

The response to the multiple-choice versions of the Cab and Light-Bulb Problems are presented in Table 4. Both problems produced similar results. The most frequent answer was a, which followed the individuating information exactly and asserted that the base rate was irrelevant. Response b, which followed the base rate, was chosen least often in both problems. Response c, which stated that both types of information were important, was given by about 37% of the subjects tested on the two problems. This

percentage can be contrasted with the 13% of the subjects whose estimates fell in the middle range (40%-60%) on problems 1 and 8 in Table 3, which were comparable except for the multiple-choice format. The difference in proportions is statistically significant ($Z = 3.12$; $p < .001$), and suggests that there may be a little more (but by no means adequate) appreciation of the importance of the base rate than is implied by the results of the direct estimation (non-multiple-choice) format. Table 4 also indicates that subjects were, on the average, moderately confident, regardless of which answer they chose.

Insert Table 4 about here

Discussion

The present study indicates that the dominance of individuating information over prior probabilities is a robust phenomenon, impervious to incidental features of the basic inference task as well as to major changes in the content of the problem. The irrelevance of content casts doubt upon Hammerton's hypothesis about the importance of the certain types of individuating information. All types were found here to be dominant. A further weakness of Hammerton's explanation is its assumption that subjects' prior expectations about specific information are learned from experience. Since the world operates according to Bayes' Theorem, experience should confirm the importance of base rates. In light of these difficulties, the representativeness hypothesis remains the most attractive general explanation of the present results.

The representativeness hypothesis predicts that probability estimates will be determined by the most salient feature of the evidence which, in these problems, is the stated accuracy of the witness or testing device.

When evidence specific to a case is known to be correct with probability p , few persons recognize that base rate considerations work to make the accuracy dependent upon which state the evidence implies. People are confident that a device which is 80% accurate regardless of the true state, is always 80% accurate. It appears to be so simple that they wonder why anyone would even question the stated probability.

One might expect a misunderstanding as pervasive as this to be observable beyond the confines of this study, in the world where professionals make decisions of importance to society. It is. Several decades after the beginning of the psychometric movement, Meehl and Rosen (1955) found it necessary to point out the fallaciousness of trying to assess the predictive validity of a psychological test on the basis of hit rate alone, without reference to base-rate considerations. Insensitivity to the effects of prior probabilities was a common shortcoming of test developers and users, as Meehl and Rosen showed.

The lesson that Meehl and Rosen taught psychometricians holds as well for any domain in which judges attempt to use their expertise to predict low base-rate phenomena. Unless the judge is extremely accurate regardless of the true state, he is more likely to be wrong than right when he predicts the rare event. The type of misunderstanding observed in the present study may help explain why people so often predict, with confidence, rare events.

Finally, the contrast between the present results and the conservatism found in other studies or probabilistic judgment (e.g., Edwards, 1968) deserves comment. In those studies, subjects typically observed samples of data drawn from one of several sources. Their task was to estimate the probability that a particular source was generating the data. These

estimates were found to stay too close to the prior probability for the source--hence the label conservatism. However, later analyses (Kahneman & Tversky, 1972; Slovic & Lichtenstein, 1971) showed that subjects' conservatism in these tasks resulted from improper operations which they performed on the sample data, rather than from overweighting of prior probabilities. Thus, the contrast between the "conservatism studies" and the present study may be more apparent than real. In both types of problems, subjects' lack of insight leads them to neglect prior probabilities and rely almost exclusively on specific evidence.

Reference Notes

1. Kahneman, D. and Tversky, A. On the psychology of prediction.

Oregon Research Institute Research Bulletin, 1972, Vol. 12, No. 4.

References

- Edwards, W. Conservatism in Human information processing. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley, 1968.
- Hammerton, M. A case of radical probability estimation. Journal of Experimental Psychology, 1973, 101, 252-254.
- Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.
- Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.
- Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 1955, 52, 194-216.
- Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Footnotes

This study was supported by the Advanced Research Projects Agency of the Department of Defense (ARPA Order No. 2449) and was monitored by ONR under Contract No. N00014-73-C-0438 (NR 197-026).

The critical comments of Robyn Dawes, Baruch Fischhoff, and Sarah Lichtenstein are greatly appreciated.

Requests for reprints should be sent to Paul Slovic, Oregon Research Institute, P. O. Box 3196, Eugene, Oregon, 97403.

Table 1
Expected Results from 10,000 Simulations
of the Hammerton Problem

		Test Says		
		+	-	
Disease State	+	90	10	100
	-	99	9801	9900

Probability of disease given positive test score = $90/(90+99) = .48$

Table 2
Expected Results from 100 Simulations
of the Cab Problem

		Witness Says		
		Blue	Green	
True Color	Blue	68	17	85
	Green	3	12	15

Probability that True Color is Blue given witness says Green = $17/(17+12) = .59$

Table 3

Results from the Cab Problem, Its Variations, and the Hammerton Problems

Problem	No. of Subjects	Base Rate Implies	Individuating Information Implies	Bayesian Answer	Median Estimate	Inter- quartile Range
1. Cab: Standard Version	31	.85 Blue	.20 Blue	.59 Blue	.20 Blue	.20-.40
2. Cab: Individuating Info. First	23	.85 Blue	.20 Blue	.59 Blue	.25 Blue	.15-.65
3. Cab: Probability Green Estimated	29	.85 Blue	.80 Green	.41 Green	.80 Green	.80-.80
4. Cab: Confirming Evidence (1)	30	.85 Blue	.60 Blue	.90 Blue	.60 Blue	.60-.75
5. Cab: Confirming Evidence (2)	35	.60 Blue	.80 Blue	.86 Blue	.80 Blue	.73-.80
6. Cab: Base Rate Alone	28	.85 Blue	----	.85 Blue	.85 Blue	.85-.85
7. Light Bulb	59	.85 Good	.20 Good	.59 Good	.20 Good	.20-.80
8. Light Bulb: Random Selection	56	.85 Good	.20 Good	.59 Good	.20 Good	.15-.59
9. Light Bulb: Numerical Base Rate	14	.80 Good	.25 Good	.57 Good	.25 Good	.25-.25
10. Light Bulb: 99:1 Base Rate; Prob. (Good) Estimated	30	.99 Good	.20 Good	.96 Good	.35 Good	.20-.78
11. Light Bulb: 99:1 Base Rate; Prob. (Defective) Estimated	57	.99 Good	.80 Def.	.04 Def.	.80 Def.	.74-.80
12. Right Hander	55	.85 R.H.	.20 R.H.	.59 R.H.	.20 R.H.	.15-.50
13. Disease (Hammerton)	28	.01 +	.90 +	.48 +	.90 +	.89-.99
14. Engine-Crack (Hammerton)	25	.01 +	.90 +	.48 +	.90 +	.81-.90

Table 4

Distribution of Responses to the Multiple-Choice Problems

<u>Cab Problem</u>			<u>Light Bulb Problem</u>		
<u>Response</u>	<u>Frequency</u>	<u>Mean Confidence</u>	<u>Response</u>	<u>Frequency</u>	<u>Mean Confidence</u>
a. (20%)	14	3.3	a. (20%)	12	3.4
b. (85%)	4	3.0	b. (85%)	3	3.0
c. (about 50%)	9	3.2	c. (about 50%)	10	3.0